

## PR2.1. ALGORITMOS DE APRENDIZAJE SUPERVISADO: NAIVE BAYES

### Descripción de la tarea

---

Realizar estudios de aprendizaje supervisado utilizando los algoritmos bayesianos. Para ello en esta tarea se utilizarán dos datasets.

1. <https://www.kaggle.com/yasserh/wine-quality-dataset> . Clasificación de la calidad del vino según alguno de sus parámetros (ph, densidad, sulfatos, alcohol, azúcar, etc.). El dataset contiene una serie de datos sobre un vino y su calidad (de 0 a 10). Este problema es de clasificación puesto que consiste en clasificar el vino en concreto en su clase de calidad (valores discretos del cero al 10)
2. <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>. Clasificación binaria. El objetivo es clasificar entre presencia de enfermedad del corazón o no a partir de una serie de atributos (edad, sexo, colesterol, etc.). Para ello el dataset contiene una serie de datos pacientes etiquetados si tienen enfermedad o no.

Para cada uno de ellos realizar un documento de google colab con los siguientes epígrafes y tareas:

- Importación de librerías necesarias
- Preproceso
  - Importación de los datos del dataset
  - Mostrar las primeras y últimas filas del dataframe importado
  - Mostrar parámetros estadísticos de los datos (media, desviación típica, cuartiles, etc.)
  - Mostrar un mapa de calor que indique la correlación entre variables
  - Seleccionar las características a tener en cuenta en el estudio
  - Separar datos entre datos de entrada y etiquetas (resultados)
  - Separar datos entre entrenamiento y prueba (usando un 75% para entrenamiento y 25% para test)
- Entrenamiento y predicción
  - Elegir, instanciar y entrenar el modelo
  - Realizar una predicción con los datos de prueba
- Evaluación

- Mostrar el porcentaje de elementos correctamente clasificados
- Mostrar la predicción realizada (imprimir la variable con la predicción)
- Representar gráficamente la clasificación obtenida (matriz de confusión)
- Optimización (probar el proceso con distintos clasificadores)
  - Finalmente prueba los distintos clasificadores y realiza una pequeña comparativa indicando cuál obtiene mejores resultados.

**Documentación a entregar.** Se entregará en moodle centros un archivo (cuaderno de júpiter o google colab) para cada uno de los ejercicios con el desarrollo de las tareas anteriormente mencionadas. Además se entregará un fichero pdf con unas conclusiones (para cada ejercicio) que describan todas las fases del proceso de aprendizaje supervisado que se han ido desarrollando en esta tarea (explicación de qué se ha realizado)

## Evaluación

---

Los criterios de evaluación de la tarea son a, b, c, d, e, f, g, h del RA3. Aplica algoritmos de aprendizaje supervisado, optimizando el resultado del modelo y minimizando los riesgos asociados.

Para evaluar la práctica se puntúan los siguientes apartados (1 punto cada uno de ellos por cada problema planteado):

- Clasificación de vinos
  - Preproceso; Entrenamiento y predicción; Evaluación; Optimización; Conclusiones.
- Clasificación presencia de enfermedad de corazón
  - Preproceso; Entrenamiento y predicción; Evaluación; Optimización; Conclusiones.

## Rúbrica por cada problema

|                                   | 1 puntos  | 0,75 puntos   | 0,5 puntos   | 0,25 puntos   | 0 puntos   |
|-----------------------------------|---|---|--|---|--|
| <b>Preproceso</b>                 | Proporciona los datos etiquetados al modelo, importándolos y realizando todas la modificaciones necesarias.               | Proporciona los datos etiquetados al modelo, hay algunos errores menores en la importación, pero los datos se adaptan                     | Proporciona los datos etiquetados al modelo, hay algunos errores en la importación o no se realizan modificaciones         | Proporciona los datos etiquetados al modelo, hay errores en la importación y no se realizan modificaciones  | No realiza la importación de datos, o la realiza de forma errónea                |
| <b>Entrenamiento y predicción</b> | Realiza el entrenamiento del modelo y la predicción con los datos de test   | Realiza el entrenamiento del modelo y la predicción con los datos de test. Pero existen algunos errores menores                           | Realiza el entrenamiento correctamente pero no la predicción   | Realiza el entrenamiento pero no la predicción y existen algunos errores en el entrenamiento  | No realiza el entrenamiento del modelo, ni la predicción o presentan errores     |
| <b>Evaluación</b>                 | Se realiza una evaluación del modelo con los datos. Además se representa de diversas formas, gráficamente en texto        | Se realiza una evaluación del modelo con los datos. Sólo se presenta un gráfico   | Se realiza una evaluación del modelo con los datos. Se muestra un gráfico pero con algunos errores menores                 | Se realiza una evaluación del modelo con los datos. Sólomente en texto  | No se realiza la evaluación o presenta errores importantes                       |
| <b>Optimización</b>               | Se realiza una optimización de hiperparámetros o una comparativa entre distintos modelos de la misma familia.             | Se realiza una optimización de hiperparámetros o una comparativa pero existen algunos errores menores o no está suficientemente detallado | Se realiza una optimización de hiperparámetros o Hay algún error en los los valores óptimos o en la comparativa.           | Se realiza una optimización de hiperparámetros o una comparativa. Existen errores importantes, no se consiguen los valores óptimos o la comparativa es escasa o irrelevante | No se realiza la optimización ni se prueban diversos modelos de la misma familia |
| <b>Conclusiones</b>               | Realiza unas conclusiones que describen todas las fases del proceso de solucionar un problema con aprendizaje supervisado | Realiza unas conclusiones que describen la mayoría de las fases del proceso de solucionar un problema con aprendizaje                     | Realiza unas conclusiones que describen al menos la optimización de parámetros y la evaluación de los resultados obtenidos | Realiza unas conclusiones que describen alguna de las fases del proceso de solucionar un problema con aprendizaje   | No realiza conclusiones o son erróneas o irrelevantes                            |

## Refuerzo y ampliación

---

### Refuerzo.

1. Realiza el estudio de aprendizaje supervisado para el problema de clasificación de los distintos tipos de flores iris según el ancho y el alto del pétalo y sépalo. Para ello utiliza el dataset iris (<https://archive-beta.ics.uci.edu/dataset/53/iris>)
2. Realiza el estudio de aprendizaje supervisado para el problema de clasificación binaria recogido en el siguiente enlace. (accidentes cerebrovasculares).  
<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>

### Ampliación.

Consulta los siguientes enlaces que contienen datasets para machine learning <https://archive-beta.ics.uci.edu/datasets> y <https://www.kaggle.com/datasets>.

Elige algún/algunos datasets y plantea y resuelve un problema de clasificación usando los algoritmos bayesianos.